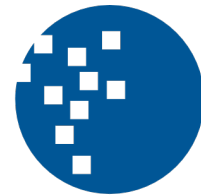


**ANALISIS KOMPARATIF MODEL MONOLINGUAL
INDOBERT DAN MODEL MULTILINGUAL XLM-R
PADA KLASIFIKASI EMOSI TEKS BAHASA
INDONESIA**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

MUHAMMAD LUTFI SALINGGIH
00000073400

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026

**ANALISIS KOMPARATIF MODEL MONOLINGUAL
INDOBERT DAN MODEL MULTILINGUAL XLM-R
PADA KLASIFIKASI EMOSI TEKS BAHASA
INDONESIA**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**MUHAMMAD LUTFI SALINGGIH
00000073400**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Muhammad Lutfi Salinggih

Nomor Induk Mahasiswa : 00000073400

Program Studi : Informatika

Skripsi dengan judul:

Analisis Komparatif Model Monolingual IndoBERT dan Model Multilingual XLM-R pada Klasifikasi Emosi Teks Bahasa Indonesia

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 8 Januari 2026



(Muhammad Lutfi Salinggih)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Muhammad Lutfi Salinggih
NIM : 00000073400
Program Studi : Informatika
Judul Laporan : Analisis Komparatif Model Monolingual IndoBERT dan
Model Multilingual XLM-R pada Klasifikasi Emosi Teks
Bahasa Indonesia

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan
Artifisial (AI) dalam pengerjaan sebagai berikut (beri tanda centang yang sesuai):

- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

1. Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
2. Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
3. Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 8 Januari 2026



(Muhammad Lutfi Salinggih)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS KOMPARATIF MODEL MONOLINGUAL INDOBERT DAN MODEL MULTILINGUAL XLM-R PADA KLASIFIKASI EMOSI TEKS BAHASA INDONESIA

oleh

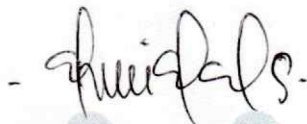
Nama : Muhammad Lutfi Salinggih
NIM : 00000073400
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 13 Januari 2026
Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

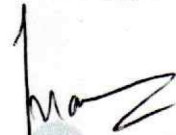
Ketua Sidang



(Eunike Endariahna Surbakti, S.Kom.,
M.T.I.)

NIDN: 0322099401

Penguji



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)

NIDN: 0818038501

Pembimbing



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Muhammad Lutfi Salinggih
NIM : 00000073400
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis Komparatif Model
Monolingual IndoBERT dan Model
Multilingual XLM-R pada Klasifikasi
Emosi Teks Bahasa Indonesia

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 8 Januari 2026

Yang menyatakan



Muhammad Lutfi Salinggih

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

”Datang buru-buru, sebelum pergi siram dulu.”

Salinggih (2025)



KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Semesta Alam Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul "Analisis Komparatif Model Monolingual IndoBERT dan Model Multilingual XLM-R pada Klasifikasi Emosi Teks Bahasa Indonesia" dengan tepat waktu.

Mengucapkan terima kasih

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga penulis yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman dan sahabat penulis yang telah membantu dan mendukung penulis, sehingga penulis bisa menyelesaikan tugas akhir ini.

Ucapan terima kasih juga penulis sampaikan kepada semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan dukungan dalam bentuk apa pun. Semoga laporan ini dapat memberikan manfaat bagi pembaca dan berkontribusi bagi pengembangan ilmu pengetahuan, khususnya di bidang kecerdasan buatan.

Tangerang, 8 Januari 2026



Muhammad Lutfi Salinggih

ANALISIS KOMPARATIF MODEL MONOLINGUAL INDOBERT DAN MODEL MULTILINGUAL XLM-R PADA KLASIFIKASI EMOSI TEKS BAHASA INDONESIA

Muhammad Lutfi Salinggih

ABSTRAK

Pertumbuhan pesat teks berbahasa Indonesia pada media sosial dan platform digital menghasilkan data yang kaya akan ekspresi emosional, namun analisis emosi masih banyak dilakukan secara manual dan subjektif sehingga tidak efisien untuk data berskala besar. Penelitian ini membandingkan kinerja model monolingual IndoBERT dan model multilingual XLM-RoBERTa dalam mengklasifikasikan lima kategori emosi, yaitu *cinta*, *gembira*, *marah*, *sedih*, dan *takut*. Dataset internal yang digunakan berjumlah 10.075 data, sedangkan dataset eksternal sebanyak 1.932 data digunakan untuk menguji kemampuan generalisasi. Kedua model dilakukan *fine-tuning* dengan konfigurasi yang identik, yaitu *learning rate* 2×10^{-5} , *batch size* 32, dan 3 *epoch*, serta dievaluasi menggunakan skema *5-fold cross-validation*. Hasil eksperimen menunjukkan bahwa IndoBERT memperoleh akurasi 89,09% pada data uji internal, sedangkan XLM-RoBERTa memperoleh akurasi 87,30%. Pada pengujian data eksternal, XLM-RoBERTa menunjukkan akurasi yang lebih stabil sebesar 86,59% dibandingkan IndoBERT sebesar 85,40%. Hasil ini menunjukkan bahwa model multilingual mampu menyamai kinerja model monolingual setelah *fine-tuning*, dengan keunggulan pada aspek generalisasi, sehingga pemilihan model sebaiknya mempertimbangkan konteks penggunaan dan karakteristik data.

Kata kunci: *Fine-tuning*, *IndoBERT*, Klasifikasi Emosi, *Transformer*, *XLM-RoBERTa*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**COMPARATIVE ANALYSIS OF INDOBERT AS MONOLINGUAL MODEL
AND XLM-R AS MULTILINGUAL MODEL FOR INDONESIAN TEXT
EMOTION CLASSIFICATION**

Muhammad Lutfi Salinggih

ABSTRACT

The rapid growth of Indonesian-language text on social media and digital platforms produces large-scale data rich in emotional expressions; however, emotion analysis is still commonly performed manually and subjectively, making it inefficient for large datasets. This study compares the performance of a monolingual model, IndoBERT, and a multilingual model, XLM-RoBERTa, in classifying five emotion categories: love, joy, anger, sadness, and fear. The internal dataset consists of 10,075 samples, while an external dataset of 1,932 samples is used to evaluate generalization. Both models are fine-tuned using identical configurations, namely a learning rate of 2×10^{-5} , batch size of 32, and 3 training epochs, and evaluated using 5-fold cross-validation. Experimental results show that IndoBERT achieves an accuracy of 89.09% on the internal test set, whereas XLM-RoBERTa achieves 87.30%. On the external dataset, XLM-RoBERTa demonstrates more stable performance with an accuracy of 86.59%, compared to 85.40% obtained by IndoBERT. These findings indicate that a multilingual model can match the performance of a monolingual model after fine-tuning, with better generalization capability, suggesting that model selection should be based on application context and data characteristics.

Keywords: *Emotion Classification, Fine-tuning, IndoBERT, Transformer, XLM-RoBERTa.*

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR KODE	xiv
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Teks dan Emosi	6
2.2 Natural Language Processing	7
2.2.1 Pra-pemrosesan Teks	7
2.2.2 Tokenisasi	7
2.3 Representasi Teks dan Word Embedding	9
2.4 Model Transformer	10
2.5 Transfer Learning dan Fine-Tuning pada Model Bahasa	12
2.6 BERT	13
2.7 IndoBERT	14
2.8 RoBERTa	14
2.9 XLM-R	16
2.10 Cross-Entropy Loss	16
2.11 Metrik Evaluasi	17
BAB 3 METODOLOGI PENELITIAN	20
3.1 Spesifikasi dan Environment Eksperimen	21
3.2 Pengumpulan Data	21
3.3 Preprocessing Data	23
3.3.1 Data Cleaning	24
3.3.2 Normalisasi Teks	24
3.3.3 Pembagian Data 70:15:15	24
3.3.4 Tokenisasi: WordPiece dan SentencePiece	25
3.4 Modelling	25
3.5 Cross Validation	25
3.6 Fine-Tuning Model	26

3.6.1	Konfigurasi Hyperparameter	26
3.6.2	Analisis Dinamika Fine-Tuning Berdasarkan Jumlah Epoch	27
3.7	Evaluasi Model	28
3.7.1	Evaluasi pada Data Uji Internal	28
3.7.2	Analisis Performa per Kelas Emosi	28
3.7.3	Uji Generalisasi Model pada Dataset Eksternal	29
BAB 4	HASIL DAN DISKUSI	30
4.1	Preprocessing Data	30
4.1.1	Data Cleaning	30
4.1.2	Normalisasi Teks	31
4.1.3	Data Splitting	33
4.1.4	Tokenisasi	34
4.2	Modelling	35
4.3	Cross Validation	36
4.4	Fine-Tuning Model	38
4.5	Evaluasi Model	39
4.5.1	Hasil Evaluasi IndoBERT	42
4.5.2	Hasil Evaluasi XLM-RoBERTa	49
4.6	Analisis Komparatif Model	56
4.6.1	Perbandingan Performa Keseluruhan pada Data Uji Internal	56
4.6.2	Perbandingan Robustness dan Kemampuan Generalisasi Model	57
4.6.3	Perbandingan Performa Antar Kelas Emosi	58
BAB 5	SIMPULAN DAN SARAN	60
5.1	Simpulan	60
5.2	Saran	60
DAFTAR PUSTAKA		62



DAFTAR TABEL

Tabel 3.1	Spesifikasi perangkat dan lingkungan eksperimen	21
Tabel 3.2	Distribusi kelas emosi pada dataset internal	22
Tabel 3.3	Distribusi kelas emosi pada dataset eksternal (Kaggle) . . .	23
Tabel 4.1	Contoh <i>tweet</i> sebelum dan sesudah <i>cleaning</i>	31
Tabel 4.2	Contoh <i>tweet</i> sebelum dan sesudah normalisasi	32
Tabel 4.3	Hasil pembagian dataset	33
Tabel 4.4	Distribusi kelas emosi pada <i>training set</i>	34
Tabel 4.5	Hasil 5-fold cross validation pada data pelatihan	37
Tabel 4.6	Hasil evaluasi baseline model IndoBERT tanpa <i>fine-tuning</i>	42
Tabel 4.7	Hasil evaluasi model IndoBERT setelah <i>fine-tuning</i> pada data uji internal	43
Tabel 4.8	Perbandingan performa model IndoBERT sebelum dan sesudah <i>fine-tuning</i> pada data uji internal	43
Tabel 4.9	Hasil evaluasi model IndoBERT pada dataset eksternal . .	46
Tabel 4.10	Ringkasan performa IndoBERT pada variasi jumlah epoch	48
Tabel 4.11	Hasil evaluasi baseline model XLM-RoBERTa tanpa <i>fine-tuning</i>	50
Tabel 4.12	Hasil evaluasi model XLM-RoBERTa setelah <i>fine-tuning</i> pada data uji internal	51
Tabel 4.13	Perbandingan performa model XLM-RoBERTa sebelum dan sesudah <i>fine-tuning</i> pada data uji internal	51
Tabel 4.14	Hasil evaluasi model XLM-RoBERTa pada dataset eksternal	53
Tabel 4.15	Ringkasan performa XLM-RoBERTa pada variasi jumlah epoch	55
Tabel 4.16	Perbandingan performa keseluruhan IndoBERT dan XLM-R pada data uji internal	57
Tabel 4.17	Perbandingan performa generalisasi IndoBERT dan XLM-R pada dataset eksternal	57
Tabel 4.18	Ringkasan hasil 5-fold cross-validation	58
Tabel 4.19	Perbandingan F1-score antar kelas emosi pada data uji internal	59

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1	Visualisasi <i>embedding</i> pada <i>BERT</i>	13
Gambar 2.2	Visualisasi <i>confusion matrix</i>	18
Gambar 3.1	Diagram alur metodologi penelitian	20
Gambar 3.2	Diagram alur preprocessing data	23
Gambar 4.1	Confusion matrix hasil fine-tuning model IndoBERT pada data uji internal	44
Gambar 4.2	Kurva akurasi dan <i>loss</i> selama proses fine-tuning model IndoBERT	45
Gambar 4.3	Confusion matrix hasil uji generalisasi model IndoBERT	47
Gambar 4.4	Confusion matrix hasil fine-tuning model XLM-RoBERTa pada data uji internal	52
Gambar 4.5	Kurva akurasi dan <i>loss</i> selama proses fine-tuning model XLM-RoBERTa	52
Gambar 4.6	Confusion matrix hasil uji generalisasi model XLM-RoBERTa	54



DAFTAR KODE

Kode 4.1	Implementasi data cleaning	30
Kode 4.2	Implementasi normalisasi teks	32
Kode 4.3	Pembagian dataset secara stratified	33
Kode 4.4	Tokenisasi pada data menggunakan AutoTokenizer	34
Kode 4.5	Implementasi pemodelan IndoBERT dan XLM-R	35
Kode 4.6	Implementasi 5-fold cross validation	36
Kode 4.7	Implementasi fine-tuning model	38
Kode 4.8	Evaluasi pada internal test set	39
Kode 4.9	Evaluasi pada external test set	40



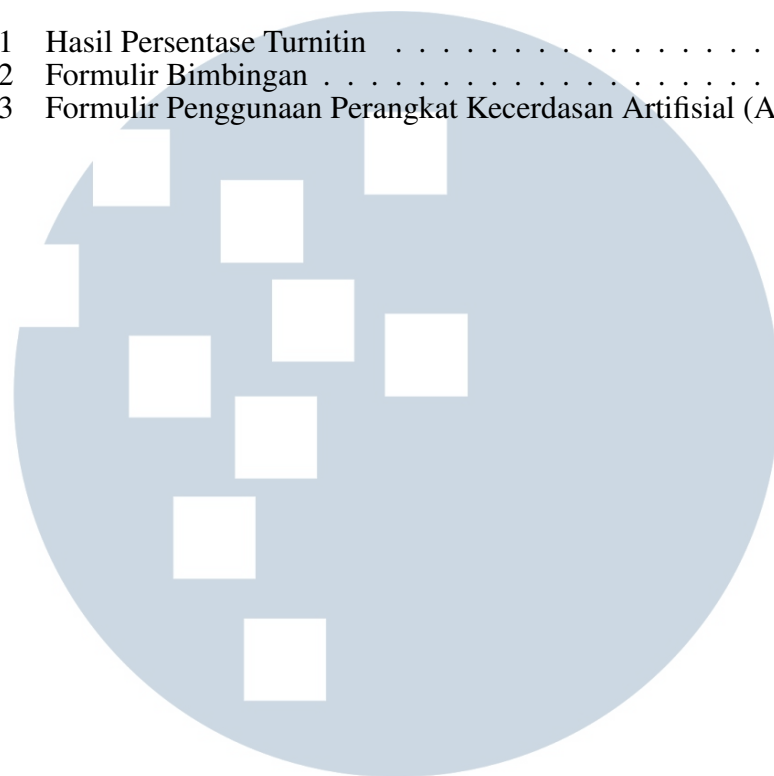
DAFTAR RUMUS

Rumus 2.1	<i>Scaled Dot-Product Attention</i>	10
Rumus 2.2	<i>Multi-Head Attention</i>	11
Rumus 2.4	<i>Feed-Forward Network</i>	11
Rumus 2.6	<i>Cross-Entropy Loss</i>	17
Rumus 2.7	<i>Akurasi</i>	18
Rumus 2.8	<i>Precision</i>	18
Rumus 2.9	<i>Recall</i>	19
Rumus 2.10	<i>F1-Score</i>	19



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	65
Lampiran 2	Formulir Bimbingan	74
Lampiran 3	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	76



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA